# Customising 3D Cell Segmentation to Study Preimplantation Mouse Embryos

Anita Karsa[1,2] Ahmed Abdelbaki[2,3,4], Kathy Niakan[2,3,4], Leila Muresan[1,2]

[1]Centre for Advanced Image Analysis (CAIC), University of Cambridge

[2]Dept. of Physiology, Development and Neuroscience (PDN), University of Cambridge

[3]Niakan Lab, The Francis Crick Institute

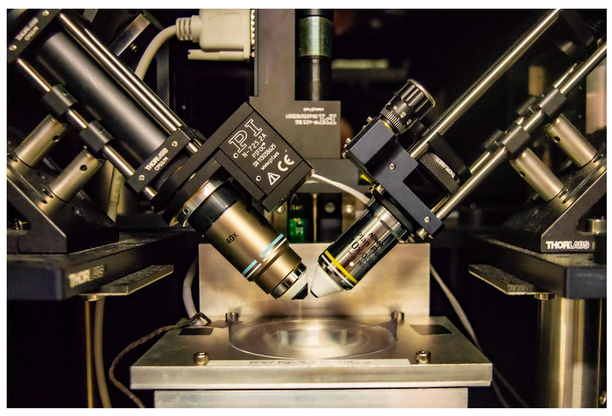
[4]Centre for Trophoblast Research, PDN, University of Cambridge

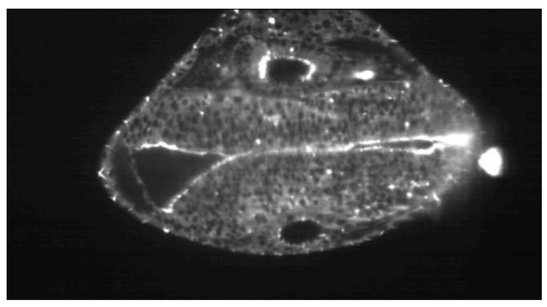
## Motivation

- Biologists hope to obtain **critical information on fertility** by studying the preimplantation development of mouse embryos[1]
- **Light sheet microscopy**[2] enables them to acquire high-resolution **3D+time images** to investigate cell division and cellular arrangement across time (**Figure 1**)
- **3D cell segmentation** is a crucial first step[3] in the quantitative analysis of these images. However, given the enormous size of these rich datasets, manual segmentation is **extremely time-consuming**
- **Deep-learning**-based cell segmentation has been shown to be very promising in 2D[4-6], but most of the pre-trained 3D neural networks **generalise very poorly** due to the lack of annotated 3D data

Here we **customised StarDist 3D**[7], a 3D U-Net-based neural network designed for cell segmentation, to provide our biologist collaborators with a **fast and accurate image segmentation pipeline** for their upcoming studies

**Figure 1**: The reduced phototoxicity of light sheet microscopy (left) leads to longer sample life which enables the acquisition of rich, 3D+time images (right)





3D+time data of zebrafish (400 GB)

Image courtesy of Dr Clare Buckley and Amelia Race

## Training StarDist 3D

- **StarDist 3D**[7] is a 3D **U-Net**-based neural network designed for cell segmentation
- We trained StarDist 3D on a set of **representative mouse embryo data** acquired by Ahmed Abdelbaki. This data had **anisotropic resolution**, i.e. the in-plane resolution was much higher than the through-plane resolution
- The training (85%) and validation (15%) data was **annotated (in 3 days)** using:
  1. the well-established **StarDist 2D**[4] **in each slice** (i.e. in plane)
  2. **manual correction** of the 2D segmentations
  3. **through-slice linear assignment particle**[8] **"tracking"** (based on in-plane distance) to generate 3D segmentations
  4. **manual correction** of the 3D segmentations
- The number of **layers, kernel sizes, and pooling sizes** in StarDist 3D **were carefully selected** to ensure an **adequate network field-of-view** for the cells
- The network was **trained (Figure 2) on the CSD3 cluster** of the University of Cambridge for 30 hours (epochs = 150, steps per epochs = 32, augmentation = random flips, and random rotations about the through-plane direction)

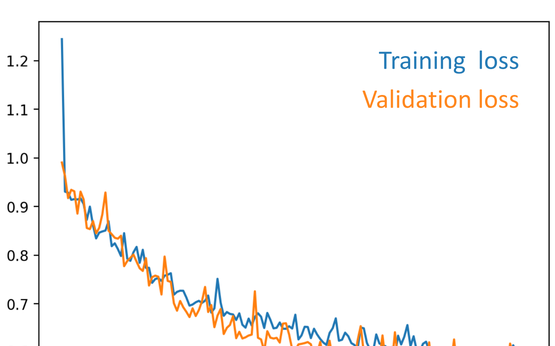



**Figure 2**: Training (blue) and validation (orange) loss across 150 epochs while training on the CSD3 cluster (4 NVIDIA Ampere A100 GPUs, with 1000GB host RAM and two AMD EPYC 64-core processors)

**References:** [1]Biggers. J Dev Biol 42.7 (2004), [2]Gibbs et al. Front Cell Dev Biol 9. (2021), [3]Nunley et al. bioRxiv (2023), [4]Schmidt et al. MICCAI Proceedings (2018), [5]Mandal et al. IEEE 18th ISBI (2021), [6]Stringer et al. Nature methods 18.1 (2021), [7]Weigert et al. In Proceedings of the IEEE/CVF WACV (2020), [8]Fukai and Kawaguchi. Bioinformatics 39.1 (2023)

## Post processing

- Despite the promising quality of the customised StarDist 3D network, additional **post processing is** still required and **crucial for accurate cell segmentation**
- **Dust particles** in the upper slices of the image often have a similar appearance to cells in light sheet microscopy images. The easiest way we found to filter these out is to manually specify a slice band that contains the edge of the embryo, delete all "cells" above, and **manually identify** dust particles in the band using an in-house Jupyter notebook (similar to **Figure 4**) at each time point **(in 5 minutes)**
- **Overlapping regions** can appear when the network identifies the same cell twice for some reason and places one on top of the other in its segmentation output (**Figure 3, black arrows**). One of the regions appears as a(n often disjointed) "halo" around the other. We **automatically identify and merge these (Figure 3, blue arrows)**

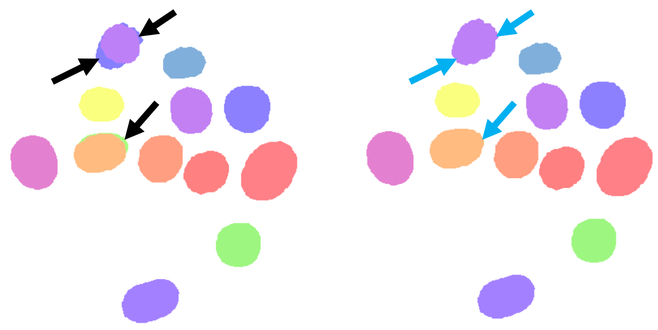
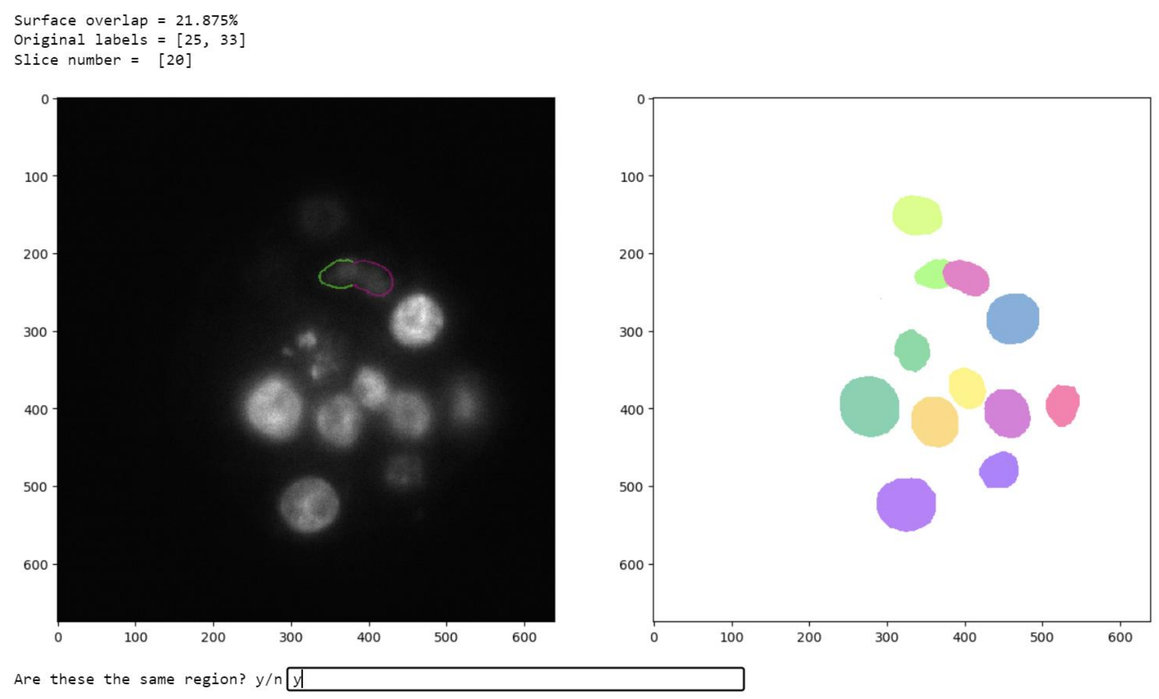


**Figure 3**: Merging overlapping labels. Some segmented regions seem to be overlapping leading to disjointed parts of one region appearing around another (black arrows). These were identified and merged (blue arrows) automatically.

- It has been reported before[5] that **StarDist 3D struggles with identifying banana-shaped regions** because it is fitting star convex shapes under the hood. Instead, it often segments these shapes as multiple regions with considerable surface overlap (Figure 4). However, there is often also a large surface overlap between cells when they are densely packed. Therefore, we perform **supervised merging (in 10 minutes per dataset) of regions with a large surface overlap** using an in-house Jupyter notebook (**Figure 4**)

**Figure 4**: In-house Jupyter notebook interface for supervised merging of regions with a large (>10%) surface overlap.




## Conclusions

- Here we **successfully customised 3D StarDist** to the data of our biologist collaborators
- All the **necessary post-processing** steps are **either fully automated or** can be **easily performed** using our in-house Jupyter notebooks
- The **overall time cost** of the new pipeline to the user is **about 15-20 minutes** to process a dataset of about 100 time points
- This is **much lower than** the amount of time we recorded during **data annotation (3 days)**

**Acknowledgements:** This work was funded by the EPSRC award EP/Y008715/1